Inequalities in the completeness of OpenStreetMap buildings in urban centers

Benjamin Herfort[1,2]*, Sven Lautenbach[1], João Porto de Albuquerque[3], Jennings Anderson[4] and Alexander Zipf[2]

[1] Heidelberg Institute for Geoinformation Technology, Heidelberg, Germany; benjamin.herfort@heigit.org, sven.lautenbach@heigit.org
[2] GIScience Chair, Institute of Geography, Heidelberg University, Heidelberg, Germany; zipf@uni-heidelberg.de
[3] Urban Big Data Centre, University of Glasgow, United Kingdom; Joao.Porto@glasgow.ac.uk
[3] Meta Platforms Inc., United States; jennings.anderson@gmail.com

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

The collaborative maps of OpenStreetMap (OSM) have become a major source of geospatial baseline data for humanitarian organisations, companies and public authorities. Describing the elements of spatial data quality (e.g. positional accuracy, completeness, temporal quality) for the OSM dataset is a key prerequisite to provide the potential stakeholders with the necessary information to decide on the fitness for use of a data set for their particular application [1]. Better spatial data quality assessment would promote the adoption and (right) usage of new sources of data such as OSM and data products based on OSM. A large community of researchers has analyzed the quality of OSM data in comparison to authoritative reference data sets, by means of remote sensing and using intrinsic measures [2–4]. It has been acknowledged that the OSM data in general is strongly biased, in part due to a much larger contributor basis in countries in the Global North as a consequence of socio-economic inequalities and the digital divide [5,6].

Albeit the manifold usage of OSM building footprints, an adequate investigation into their completeness on the global scale has not been conducted so far. This talk investigates OSM building completeness in regions home to a population of 3.5 billion people (about 50% of the global population). First, we propose a machine learning regression method based on Random Forests (RF) to assess OSM building completeness within all 13,189 urban centers (as defined by the European Commission [7]). The analysis utilizes an extensive collection of open building data from commercial and authoritative sources as training data and builds upon very recent technological advances to utilize OSM full-history data for spatio-temporal data analysis on the global scale [8]. The model further relies on information obtained from remote sensing data (land cover, population distribution, night time lights), subnational human development, and urban road network density as predictors. This allows us – for the first time – to present a comprehensive assessment of the evolution of urban OSM building completeness which encompasses all data contributed to OSM since 2008.

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2022). Inequalities in the completeness of OpenStreetMap buildings in urban centers.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004534

For each urban center we calculated the OSM building completeness using the area ratio method which has been applied by several other researchers in the context of urban areas [9–11]. Several measures have been adopted to describe the temporal evolution of inequality in urban OSM building mapping on the global scale and per World Bank region. We report on the average monthly OSM building completeness for urban centers globally and distinguished this score further by World Bank region and SHDI. These analyses has been conducted for annual snapshots from 2008-01-01 up until 2022-01-01.

We investigated the impact of humanitarian mapping through the HOT Tasking Manager and corporate mapping by Apple, Meta, MapBox, Microsoft and Kaart on overall completeness and inequality measures. OSM contributions have been considered as humanitarian mapping activities following the approach developed by Herfort et al. which utilizes information obtained from a HOT Tasking Manager database dump [12]. Corporate mapping activities were identified by OSM user ID, expanding on the approach presented in [13] by using a mapper's self-disclosed corporate affiliation in their OSM user bio instead of relying on out-of-date lists on the OSM wiki [14]. Based on this information, we derived the share of humanitarian map edits and corporate map edits on the overall OSM building data.

Overall, average urban OSM building completeness is estimated at 21% globally. Relatively high completeness was estimated for Europe & Central Asia (67%) as well as for North America (56%). Completeness values lower than the global average were observed for the regions Latin America & Caribbean (17%), East Asia & Pacific (16%), Middle East & North Africa (11%), and South Asia (7%). The completeness value for East Asia & Pacific was strongly influenced by the fact that urban centers in China were hardly mapped, very likely because mapping in OSM is prohibited by law. Sub-Saharan Africa completeness (29%) was slightly higher than the global mean. These regional differences illustrate that the global average is of limited explanatory power.

Distinguishing urban centers by SHDI also revealed dramatic differences in the temporal trajectories of completeness. In general, urban centers in regions with very high SHDI had the highest levels of mapped building completeness. Surprisingly, however, there was no positive correlation between SHDI and completeness. The completeness in low SHDI urban centers was higher than the completeness of urban centers with high SHDI. Our results suggest that this was due to the positive impact of organized humanitarian mapping activities since 2015, especially on urban centers located in low and medium SHDI regions.

We found that organized humanitarian mapping activities in urban centers contributed an average of about 8% of the building footprints globally. However, humanitarian contributions were focused on specific regions, especially in Africa where about 43% of all building edits in Sub-Saharan Africa were related to organized humanitarian mapping activities. Overall, organized humanitarian mapping activities were expectedly associated with lower subnational human development index values, in line with previous findings [12]. We generally found corporate mapping activity to constitute less than 2% of all building edits globally (and only about 0.1% in urban centers), a significant difference in participation from corporate mappers editing nearly 20% of the global road network as previously found [13].

The results reveal that for 1,510 cities home to a population of more than 400 Million people, OSM building footprint data is more than 80% complete and can provide an alternative to otherwise complex approaches utilized to derive authoritative and/or automated building datasets. The digital divide in OSM has receded over the past decade, but still exists. As such, OSM data completeness improved, but was still strongly biased by regional, socio-economic and demographic factors on several scales. This echoes the highly uneven geographies of participation observed in Wikipedia [15] and stands in contrast to the relatively higher and more evenly distributed completeness for OSM's road network [16]. If this trend continues, OSM will become more complete, but will still not evolve towards a truly global inclusive map. As a consequence, global studies and global frameworks (such as SDGs) which use OSM data will draw wrong conclusions and will provide misleading recommendations for decision makers when the biases in OSM's coverage are not accounted for.

The results reveal the need to address the remaining stark data inequalities, which could not be turned around so far by humanitarian and corporate organized mapping activities. We conclude with recommendations directed at stakeholders working with OSM data: (1) Multi-scale building completeness measures should be applied before subsequent usage of OSM data to outline the potential negative effect of missing data. (2) Completeness maps should be used in combination with socio-demographic information to guide future mapping activities to ensure that "nobody is left behind" as encouraged by the SDGs.

All Python code and Jupyter notebooks necessary to calculate the geospatial statistics, create maps and derive figures are available in this GitHub repository: https://github.com/GIScience/global-urban-building-completeness-analysis.

References

[1] Oort, P. (2006). *Spatial data quality: from description to application*. PhD thesis, Wageningen Universiteit.

[2] Neis, P., Zielstra, D. & Zipf, A. (2011). The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet, 4,* 1–21.

[3] Barron, C., Neis, P. & Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions GIS, 18*, 877–895.

[4] Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C. & Haklay, M. (2016). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science, 8816*, 1–29.

[5] Sui, D., Goodchild, M. & Elwood, S. (2013). Volunteered geographic information, the exaflood, and the growing digital divide. In: Sui, D., Elwood, S., & Goodchild, M. (Eds.) *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice,* Springer, Dordrecht, 1–12.

[6] Neis, P., Zielstra, D. & Zipf, A. (2013). Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet, 5,* 282–300.

[7] Florczyk, A., Corbane, C,. Schiavina, M., Pesaresi, M., Maffenini, L., Melchiorri, M., Politis, P., Sabo, F., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., & Zanchetta, L. (2019). GHS Urban Centre Database 2015, multitemporal and multidimensional attributes. *European Commission, Joint Research Centre*. Retrieved from https://ghsl.jrc.ec.europa.eu/ghs_stat_ucdb2015mt_r2019a.php

[8] Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.B., & Zipf, A. (2019). OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, *4*(1), 1–12.

[9] Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science, 28*, 700–719.

[10] Törnros, T., Dorn, H., Hahmann, S. & Zipf, A. (2015). Uncertainties of completeness measures in OpenStreetMap–A case study for buildings in a medium-sized German city. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W5*, 353–357.

[11] Hecht, R., Kunze, C. & Hahmann, S. (2013). Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS International Journal of Geo-Information, 2,* 1066–1091.
[12] Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J. & Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports, 11*(1), 1-15.
[13] Anderson, J., Sarkar, D. & Palen, L. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information, 8*, 232.
[14] OpenStreetMap Wiki (2022). Organised Editing/Activities. Retrieved from https://wiki.openstreetmap.org/wiki/Organised_Editing/Activities
[15] Graham, M., Straumann, R.K. & Hogan, B. (2015). Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia. *Annals of the American Association of Geographers, 105,* 1158–1178.
[16] Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PLOS ONE, 12*(8), e0180698.